

## PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2002-251234

(43)Date of publication of application : 06.09.2002

(51)Int.Cl.

G06F 3/00

G06F 3/16

G06T 7/00

G06T 7/20

G10L 15/00

(21)Application number : 2001-048656

(71)Applicant : FUJITSU LTD

(22)Date of filing : 23.02.2001

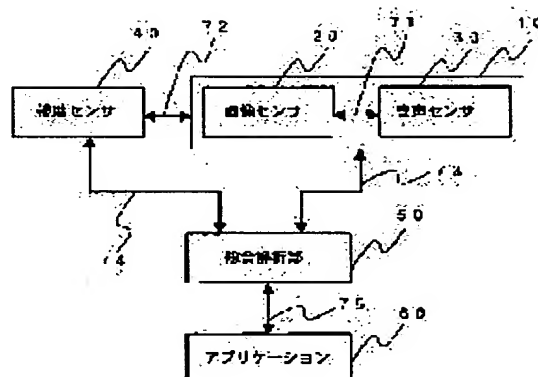
(72)Inventor : WATANABE MASANORI  
MURASE KENTARO  
NODA TAKUYA  
WATANABE KAZUHIRO

## (54) HUMAN INTERFACE SYSTEM BY PLURAL SENSOR

## (57)Abstract:

PROBLEM TO BE SOLVED: To prevent the malfunction of a man/machine I/F, and to improve the recognition rate or processing speed by selecting proper information as the input information of the man/machine I/F.

SOLUTION: A picture sensor 20, a voice sensor 30, an auxiliary sensor part 40(infrared sensor or the like), an integral analyzing part 50, and an AP 60 are allowed to communicate with each other by using data/control signal communicating means 71-75. Each sensor feeds back its own signal detection result and control information to be used by another sensor for deciding the range and detection precision of an object to be detected at the time of acquiring a signal the next time through the communicating means to another sensor. The integral analyzing part 50 checks whether or not any contradiction is generated among the detected results of the respective sensors, and applies the control information to the respective sensors. The respective sensors 20-40 decide the range and detection precision of the object of signal detection based on the acquired information, and acquire the signal according to the decision.



(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開2002-251234

(P2002-251234A)

(43)公開日 平成14年9月6日(2002.9.6)

(51)Int.Cl. <sup>7</sup>	識別記号	F I	テーマコード(参考)
G 0 6 F 3/00	6 0 1	G 0 6 F 3/00	6 0 1 5 D 0 1 5
	6 8 0		6 8 0 C 5 E 5 0 1
3/16	3 4 0	3/16	3 4 0 A 5 L 0 9 6
G 0 6 T 7/00		G 0 6 T 7/00	P
7/20	3 0 0	7/20	3 0 0 A

審査請求 未請求 請求項の数5 O L (全 22 頁) 最終頁に続く

(21)出願番号 特願2001-48656(P2001-48656)

(22)出願日 平成13年2月23日(2001.2.23)

(71)出願人 000005223

富士通株式会社

神奈川県川崎市中原区上小田中4丁目1番  
1号

(72)発明者 渡辺 正規

神奈川県川崎市中原区上小田中4丁目1番  
1号 富士通株式会社内

(72)発明者 村瀬 健太郎

神奈川県川崎市中原区上小田中4丁目1番  
1号 富士通株式会社内

(74)代理人 100095555

弁理士 池内 寛幸

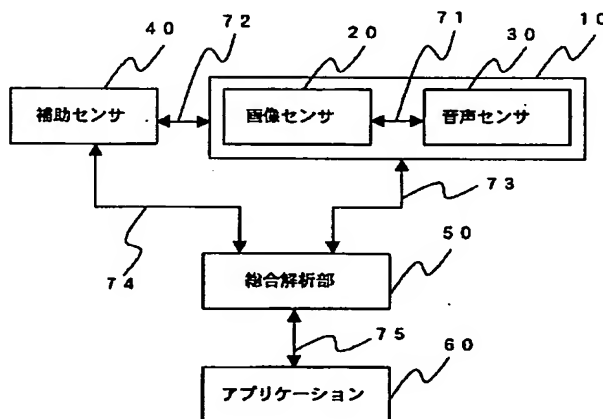
最終頁に続く

(54)【発明の名称】 複数のセンサによるヒューマンインタフェースシステム

(57)【要約】

【課題】 マンマシンI/Fの入力情報として適切な情報を取捨選択し、マンマシンI/Fの誤動作を防止したり、認識率、処理速度の向上を図る。

【解決手段】 画像センサ20、音声センサ30、補助センサ部40（赤外線センサなど）、総合解析部50、AP60は、データ／制御信号通信手段71～75により相互に通信し合う。各センサが、前記通信手段を介して、他のセンサに対して、自身の信号検出結果と、他のセンサが次の信号獲得時における検出対象の範囲と検出精度の決定に利用する制御情報をフィードバックする。総合解析部50は、各センサで検出された結果間に相互に矛盾がないか否かを調べ、各センサに制御情報を与える。各センサ20～40は、得られている情報をもとに信号検出対象の範囲と検出精度を決定し、当該決定に従って信号を獲得する。



**【特許請求の範囲】**

【請求項 1】 検出対象の範囲と検出精度を決定し、当該検出精度により当該検出対象から特定の検出信号を獲得するセンサとして、検出信号が異なる少なくとも 2 種類以上のセンサと、

前記各センサで検出された結果間に相互に矛盾がないかを調べ、各センサに対する制御情報を生成する総合解析部と、

前記各センサによる検出結果を利用するアプリケーションと、

前記各センサ相互間、前記各センサと前記総合解析部間、前記総合解析部と前記アプリケーション間においてデータと制御情報を通信し合う通信手段を備え、

前記各センサが、他のセンサから得た前記信号検出結果または制御情報、前記総合解析部から得た前記制御情報のいずれかまたはその組み合わせを利用して、次の信号獲得時における検出対象の範囲と検出精度を決定し、前記各センサが、前記通信手段を介して、他のセンサに対して、自身の信号検出結果と、他のセンサが次の信号獲得時における検出対象の範囲と検出精度の決定に利用する制御情報を出力し、

前記総合解析部が、前記通信手段を介して、前記各センサが次の信号獲得時における検出対象の範囲と検出精度の決定に利用する制御情報を出力することを特徴とする複数のセンサによるヒューマンインタフェースシステム。

【請求項 2】 前記検出対象が人であり、前記センサが少なくとも画像センサと音声センサを含み、

前記画像センサの検出信号が人の画像認識情報であり、前記画像センサが、画像認識結果に基づいて検出対象の動作を解釈し、ジェスチャ入力によるコマンド入力を認識する動作認識部を備え、

前記音声センサの検出信号が人の音声認識情報であり、前記音声センサが、音声認識結果に基づいて検出対象の音声进行を解釈し、音声入力によるコマンド入力を認識する音声認識部を備えた請求項 1 に記載の複数のセンサによるヒューマンインタフェースシステム。

【請求項 3】 前記検出対象が人であり、前記センサが少なくとも画像センサと補助センサを含み、

前記画像センサの検出信号が人の画像認識情報であり、前記画像センサが、画像認識結果に基づいて検出対象の動作を解釈し、ジェスチャ入力によるコマンド入力を認識する動作認識部を備え、

前記補助センサの検出信号が人の位置情報を検出するために有効な情報である請求項 1 に記載の複数のセンサによるヒューマンインタフェースシステム。

【請求項 4】 前記検出対象が人であり、前記センサが少なくとも音声センサと補助センサを含み、

前記音声センサの検出信号が人の音声認識情報であり、前記音声センサが、音声認識結果に基づいて検出対象の

音声を解釈し、音声入力によるコマンド入力を認識する音声認識部を備え、

前記補助センサの検出信号が人の位置情報を検出するために有効な情報である請求項 1 に記載の複数のセンサによるヒューマンインタフェースシステム。

【請求項 5】 前記検出対象が人であり、前記センサが少なくとも画像センサと音声センサと補助センサを含み、

前記画像センサの検出信号が人の画像認識情報であり、前記画像センサが、画像認識結果に基づいて検出対象の動作を解釈し、ジェスチャ入力によるコマンド入力を認識する動作認識部を備え、

前記音声センサの検出信号が人の音声認識情報であり、前記音声センサが、音声認識結果に基づいて検出対象の音声を解釈し、音声入力によるコマンド入力を認識する音声認識部を備え、

前記補助センサの検出信号が人の位置情報を検出するために有効な情報である請求項 1 に記載の複数のセンサによるヒューマンインタフェースシステム。

**【発明の詳細な説明】****【0001】**

【発明の属する技術分野】本発明は、マンマシンインタフェースの技術分野に関し、特にテレビジョンやビデオをはじめとした家電製品、コンピュータなどの操作を、リモートコントローラ、マウス、キーボードなどのボタン操作による入力装置を使わず、声や身振り、手振りで行う装置に関する。

**【0002】**

【従来の技術】現在、テレビジョンやビデオをはじめとした家電製品、コンピュータなどの操作は、リモートコントローラ、マウス、キーボードなどのボタン操作による入力装置が広く用いられている。ボタン操作による入力装置を使わず、声や身振り、手振りで行う装置も開発されつつある。マンマシンインタフェースにおける人物識別やコマンド入力を、利用者の声や身振り、手振りで行う技術として特開2000-326274号公報に記載の技術が知られている。

【0003】特開2000-326274号公報に記載の技術では、人物を識別するための視覚情報を複数のカメラの撮影による得るが、被写体位置の探索制御などの撮影制御は、もっぱら当該カメラから得られる情報のみを用いてその制御を行っていた。また、音声認識に用いる音声情報も複数のマイクロフォンから得るが、音声方向の探索制御など音声入力制御は、もっぱらロボットの前左右に設置された複数のマイクから得られる情報のみを用いてその制御を行っていた。

【0004】音声入力制御については、特開平1-195499号公報に記載の技術も知られている。特開平1-195499号公報記載の技術は、セキュリティドアのように、超音波センサによる物体検出結果とカメラから取り込んだ画像

データに基づいて入室者の口の位置を見つけ、その口の方向にマイクロフォンの指向性を合わせる技術である。

【0005】

【発明が解決しようとする課題】しかし、上記従来技術には以下の問題があった。

【0006】従来の特開2000-32674号公報に記載の技術においては、装置またはロボットの目にあたるカメラからの撮影情報、耳にあたるマイクロフォンから得られる音声情報は利用されるものの、両者は独立して扱われている。図10に示した特開2000-32674号公報において開示されているブロック図においても、画像情報処理、音声情報処理相互に情報がやりとりされることは何ら示されていない。そのため、特開2000-32674号公報に記載の技術では、人間ではない人の顔を映した写真やマネキン人形を人間として認識してしまったり、人間ではない音響機器のスピーカーから流れ出た音声情報を認識してしまうという問題が起こる。このようなことはマンマシンインタフェースの意図としないどころか、人の写真、マネキン人形、人間の声以外の雑音は画像認識、音声認識にとってはノイズとなり、認識率の低下を招いたり、また、不要な対象物から得られる画像情報や音声情報を入力するため余計な情報処理が行われ、処理速度の低下を引き起こす。

【0007】次に、特開平1-195499号公報に記載の技術においては、図11のように、マイクロフォンの指向性制御において超音波センサとカメラによる被探索物の位置情報が利用されるが、音声情報の処理結果を利用することはない。また、超音波センサとカメラによる被探索物の位置検出制御においてマイクロフォンの音声情報の処理結果が利用されることもない。特開平1-195499号公報に記載の技術では、超音波センサやカメラが物体検出のためにセンサや撮影を行っているエリア（例えば、入退室のドア位置）に人物が入った場合には、マイクロフォンの指向性を合わせて音声効率よく得ることができるが、入退室管理のドア位置のようにあらかじめ狭い探索エリアが設定されている場合のみに有効な技術であって、一般にそのような限定された探索エリアがない場合では、人物が超音波センサやカメラから外れたところに立ち、音声によるコマンド入力を行う場合も多く想定され、特開平1-195499号公報に記載の技術では柔軟に対処できない。

【0008】本発明は、以上の状況を考慮し、マンマシンインタフェースの入力情報として適切な情報を取捨選択し、マンマシンインタフェースの誤動作を防止したり、認識率、処理速度の向上を図ることを目的とする。

【0009】

【課題を解決するための手段】上記課題を解決するために、本発明の複数のセンサによるヒューマンマシンインタフェースシステムは、検出対象の範囲と検出精度を決定し、当該検出精度により当該検出対象から特定の検出

信号を獲得するセンサとして、検出信号が異なる少なくとも2種類以上のセンサと、前記各センサで検出された結果間に相互に矛盾がないかを調べ、各センサに対する制御情報を生成する総合解析部と、前記各センサによる検出結果を利用するアプリケーションと、前記各センサ相互間、前記各センサと前記総合解析部間、前記総合解析部と前記アプリケーション間においてデータと制御情報を通信し合う通信手段を備え、前記各センサが、他のセンサから得た前記信号検出結果または制御情報、前記総合解析部から得た前記制御情報のいずれかまたはその組み合わせを利用して、次の信号獲得時における検出対象の範囲と検出精度を決定し、前記各センサが、前記通信手段を介して、他のセンサに対して、自身の信号検出結果と、他のセンサが次の信号獲得時における検出対象の範囲と検出精度の決定に利用する制御情報を出力し、前記総合解析部が、前記通信手段を介して、前記各センサが次の信号獲得時における検出対象の範囲と検出精度の決定に利用する制御情報を出力することを特徴とする。

【0010】上記構成により、複数異種のセンサの認識結果を相互参照し合い、相互に矛盾の生じないような信号獲得制御を行うことができ、より正確な利用者のコマンド入力の認識が可能であり、優れたヒューマンインタフェースシステムを提供することができる。

【0011】また、上記構成において、前記検出対象が人であり、前記センサが少なくとも画像センサと音声センサと補助センサを含み、前記画像センサの検出信号が人の画像認識情報であり、前記画像センサが、画像認識結果に基づいて検出対象の動作を解釈し、ジェスチャ入力によるコマンド入力を認識する動作認識部を備え、前記音声センサの検出信号が人の音声認識情報であり、前記音声センサが、音声認識結果に基づいて検出対象の音声解釈し、音声入力によるコマンド入力を認識する音声認識部を備え、前記補助センサの検出信号が人の位置情報を検出するに有効な情報であることが好ましい。

【0012】上記構成により、画像センサによる動作認識結果と、音声センサによる音声認識結果と、他のセンサ（補助センサ）により検出した人物位置情報の結果とを参考にし、相互に矛盾のない、より正確な利用者のコマンド入力の認識が可能であり、優れたヒューマンインタフェースシステムを提供することができる。

【0013】なお、上記のように、画像センサを用いた動作認識、音声センサを用いた音声認識、他のセンサを用いた人物位置情報のすべてを組み合わせた構成のみならず、センサおよび認識結果の組み合わせとして、画像センサを用いた動作認識結果と音声センサを用いた音声認識結果の組み合わせ、画像センサを用いた動作認識結果と他のセンサを用いた人物位置検出結果の組み合わせ、音声センサを用いた音声認識結果と他のセンサを用いた人物位置検出結果の組み合わせなど多様な組み合わせ

せが可能である。

【0014】

【発明の実施の形態】（実施形態 1）本発明の実施形態 1 にかかる複数のセンサによるヒューマンインタフェースシステムを述べる。

【0015】本発明の実施形態 1 の複数のセンサによるヒューマンインタフェースシステムは、カメラによる画像の撮影過程およびその認識過程、また、マイクによる音声情報の録音過程およびその認識過程において、それぞれ、それぞれが処理した過去の結果を参考するほか、互いの結果を参考する。さらに、必要に応じて、超音波センサ、赤外センサ、レンジセンサなどの補助センサを用いて得られた情報とその処理結果、また、アプリケーションの実行結果も参考する。各センサは、相互に解析結果や信号獲得制御に関する指示のやり取りを行い、さらに、総合解析部による解析結果や、各センサにおける信号獲得の制御に関する指示のやり取りを行い、センサ間の矛盾を調べつつフィードバック処理を実行し、信号を獲得する範囲および精度を確実に決定し、複数のセンサ全体で整合性のとれた出力を生成するものである。

【0016】例えば、認識処理開始当初は、各センサが環境全体から信号を獲得し、各センサにおける獲得信号を解析し、その結果を他のセンサに渡し、次の信号獲得ステップでは、前ステップにおける自らの獲得信号の解析結果、他のセンサからの解析結果、総合解析部による各センサ間で矛盾が生じないように処理した解析結果を考慮し、環境中において、信号を獲得すべき対象の範囲と精度を決定し、信号を獲得するものである。

【0017】まず、実施形態 1 にかかる複数のセンサによるヒューマンインタフェースシステムの装置構成概略を示す。

【0018】図 1 は、本発明の実施形態 1 にかかる複数のセンサによるヒューマンインタフェースシステムの装置構成概略を示す図である。10 はセンサ部で画像センサ 20、音声センサ 30 を備えている。40 は補助センサ部、50 は総合解析部、60 はアプリケーションである。ここで、補助センサ部 40 は、本発明のヒューマンインタフェースシステムが用いられる対象に応じて、画像センサ 20、音声センサ 30 以外に有用な信号を得ることができる一又は複数のセンサを備えたものである。

【0019】図 1 に示すように、本発明の実施形態 1 にかかる複数のセンサによるヒューマンインタフェースシステムでは、センサ部 10 の画像センサ 20、音声センサ 30、補助センサ部 40 は、各センサの信号処理において他のセンサで得られた信号処理結果を相互に利用し合い、かつ、他のセンサの制御を指示するものとなっている。図 1 に示すように、センサ部 10 の画像センサ 20 と音声センサ 30 間には、双方向に信号処理結果や制御データをやりとりするデータ／制御信号通信手段 71 が設けられている。また、センサ部 10 と補助センサ部

40 間にも双方向に信号処理結果や制御データをやりとりするデータ／制御信号通信手段 72 が設けられている。さらに、センサ部 10 と補助センサ部 40 と総合解析部 50 の間にも双方向に信号処理結果や制御データをやりとりするデータ／制御信号通信手段 73、74 が設けられている。総合解析部 50 とアプリケーション 60 の間にも双方向に信号処理結果や制御データをやりとりするデータ／制御信号通信手段 75 が設けられている構成となっている。

【0020】従来技術で説明した図 10 の構成では、各センサにおける信号処理は個別に実行され、それぞれ独立した情報としてアプリケーションに至るものであり、図 11 の構成では、音声センサの処理において超音波センサと画像センサの処理結果を利用するものの、情報や制御の流れが一方的であった。それに対し、本発明の構成は、センサ部 10 や補助センサ部 40 の各センサが、データ／制御信号通信手段 71 および 72 を通じて、互いに他のセンサの処理結果を利用し、また、他のセンサに対する制御を指示する。さらに、データ／制御信号通信手段 73 および 74 を通じて、総合解析部 50 の処理結果をフィードバックして利用したり、総合解析部 50 から各センサに対する制御の指示を受ける構成となっている。総合解析部 50 はセンサ部 10、補助センサ部 40 にあるすべてのセンサの処理結果を入力とし、各センサ間の矛盾を調べつつフィードバック処理を実行し、装置のセンサ全体で整合性のとれた出力を生成するものとなっている。さらに、総合解析部 50 は状況に応じて、データ／制御信号通信手段 75 を通じて、アプリケーション 60 の実行結果を参考したり、アプリケーション 60 から信号処理に関する指示を受け、それをもとに各センサを制御する構成となっている。

【0021】図 1 の基本構成を適用し、図 1 の画像センサ 20、音声センサ 30、補助センサ 40 構成例を詳しく示した図を図 2 に示す。

【0022】センサ部 10 は、画像センサ 20、音声センサ 30 を備え、画像センサ 20、音声センサ 30 は、以下のように、それぞれ、センサ制御部、センサ装置、センサ信号解析部、通信部の 4 つの部分で構成されている。

【0023】画像センサ 20 は、撮影装置 21、撮影装置制御部 22、画像解析部 23、通信部 24 の 4 つの部分から成る。

【0024】音声センサ 30 は、録音装置 31、録音装置制御部 32、音声解析部 33、通信部 34 の 4 つの部分で構成されている。

【0025】補助センサ部 40 は、本発明のヒューマンインタフェースシステムが用いられる対象に応じて、画像センサ、音声センサ以外に有用な信号を得ることができる一又は異種複数の各種センサを備えたものである。例えば、超音波センサ、赤外線センサ、レンジセンサなどが挙げられる。補助センサ部 40 に含まれる各補助セ

ンサは、それぞれ、補助センサ装置 4 1、補助センサ装置制御部 4 2、信号解析部 4 3、通信部 4 4 の 4 つの部分を備えている。

【0026】撮影装置 2 1、録音装置 3 1、補助センサ装置 4 1 は、環境中から信号を獲得する装置である。撮影装置 2 1 は、環境を撮影し、画像情報を生成する部分である。録音装置 3 1 は、録音装置環境の音響を録音し、音声情報を生成する部分である。補助センサ装置 4 1 は、環境の状態を表す信号を獲得する部分である。

【0027】通信部 2 4、3 4、4 4 は、データ／制御信号通信手段 7 1～7 4 を通じて、他のセンサまたは総合解析部 5 0 との通信インタフェースを制御する部分である。相互に解析結果や信号獲得制御に関する指示のやり取りを制御する。

【0028】撮影制御部 2 2 および画像解析部 2 3、録音装置制御部 3 2、音声解析部 3 3、補助センサ装置制御部 4 2 および信号解析部 4 3 は、それぞれ連動してセンサ信号獲得制御および信号解析を行う。センサ信号獲得制御および信号解析には、以下の、第 1 のセンサ信号獲得制御および信号解析、第 2 のセンサ信号獲得制御および信号解析、第 3 のセンサ信号獲得制御および信号解析がある。

【0029】第 1 のセンサ信号獲得制御および信号解析は、センサ信号獲得制御として、環境中において、信号を獲得する対象の範囲および獲得する精度を決定し、センサ装置（画像センサ装置 2 1、音声センサ装置 3 1、補助センサ装置 4 1）を制御する。その際、必要に応じて、データ／制御信号通信手段 7 1～7 4 を通じて、他のセンサ（画像センサ 2 0、音声センサ 3 0、補助センサ部 4 0 に含まれる他のセンサ）の解析結果、総合解析部 5 0 の結果、当該センサ自身のそれまでの解析結果を得て、それらを参考にしてセンサ装置を制御する。信号解析として、獲得した信号の処理にあたり、必要に応じて、他のセンサの解析結果、総合解析部 5 0 の結果、当該センサ自身のそれまでの解析結果を得て、それらを参考して信号を解析する。

【0030】第 2 のセンサ信号獲得制御および信号解析は、第 1 のセンサ信号獲得制御および信号解析よりさらに、他のセンサや総合解析部 5 0 との間の解析結果や制御指示のフィードバックを重視するものである。まず、センサ信号獲得制御として、データ／制御信号通信手段 7 1～7 4 を通じて、他のセンサまたは総合解析部 5 0 から受けた、環境中において、信号を獲得する対象の範囲および獲得する精度に関する指示を得て、その指示にしたがってセンサ装置を制御する。制御開始当初においてかかる指示が得られていない場合、環境中、信号を獲得する対象の範囲および獲得する精度の決定にあたり、必要に応じて、データ／制御信号通信手段 7 1～7 4 を通じて、他のセンサの解析結果、総合解析部 5 0 の結果、当該センサ装置自身のそれまでの解析結果を得て、

それらを参考にして決定し、センサ装置を制御する。信号解析として、獲得した信号の処理にあたり、必要があれば、データ／制御信号通信手段 7 1～7 4 を通じて、他のセンサの解析結果、総合解析部 5 0 の結果、当該センサ自身のそれまでの解析結果を得て、それらを参考にして信号を解析する。また、解析結果に基づき、必要に応じて、データ／制御信号通信手段 7 1～7 4 を通じて、他のセンサに対して、環境中において、信号を獲得する対象の範囲および獲得する精度に関して指示データを与える。

【0031】第 3 のセンサ信号獲得制御および信号解析は、第 2 のセンサ信号獲得制御および信号解析と同様、他のセンサや総合解析部 5 0 との間の解析結果や制御指示のフィードバックを重視しつつ、かつ、フィードバックを受動的に受けるのみでなく、フィードバック情報を採用するか否かを自律的に判断するものである。まず、センサ信号獲得制御として、データ／制御信号通信手段 7 1～7 4 を通じて、他のセンサまたは総合解析部 5 0 から受けた、環境中、信号を獲得する範囲および精度に関するフィードバックされた指示を得て、その指示に従うか否かを他のセンサの解析結果、総合解析部 5 0 の結果、センサ自身のそれまでの解析結果を参考にして判断し、従う場合は、同指示に従ってセンサ装置を制御する。従わないと判断した場合、または、他のセンサ、総合解析部 5 0 から信号を獲得する範囲および精度に関する指示のフィードバックがない場合は、それまでに得られている他のセンサの解析結果、総合解析部 5 0 の結果、センサ自身の解析結果を参考にして信号を獲得する範囲および精度を判断し、センサ装置を制御する。なお、信号解析については第 2 のセンサ信号獲得制御および信号解析で述べた信号解析と同様で良い。

【0032】次に、総合解析部 5 0 を説明する。

【0033】総合解析部 5 0 は、センサ部 1 0、補助センサ部 4 0 の解析結果、アプリケーションの実行結果を参考して、複数のセンサ間で最も矛盾の少ない結果を解析して求め、当該解析結果に基づき、必要に応じて各センサ部 1 0、補助センサ部 4 0 に環境中、信号を獲得する対象の範囲および獲得する精度に関して指示する部分である。

【0034】総合解析部 5 0 は、データ／制御信号通信手段 7 3、7 4 を通じて、センサ部 1 0、補助センサ部 4 0 の解析結果を獲得し、さらに、データ／制御信号通信手段 7 5 を通じて、アプリケーション 6 0 からセンサ制御に関する指示を得る。信号処理結果を、データ／制御信号通信手段 7 5 を通じて、アプリケーション 6 0 に提供する。さらに、必要に応じてアプリケーション 6 0 から実行結果のフィードバックを獲得する。それら得た情報をもとに、複数のセンサ間で最も矛盾の少ない結果を求める。当該解析結果に基づき、必要があれば、データ／制御信号通信手段 7 3、7 4 を通じて、センサ部 1

0、補助センサ部40に対して、環境中、信号を獲得する対象の範囲および獲得する精度に関する指示を与える。

【0035】次に、本発明の実施形態1の複数のセンサによるヒューマンインタフェースシステムの信号処理の流れを図3のフローチャートを参照しつつ説明する。

【0036】まず、センサ部10、補助センサ部40の各センサと、総合解析部50を並行して起動する(ステップS301)。

【0037】センサ部10、補助センサ部40の各センサは、撮影装置21、録音装置31、補助センサ装置41を用いて、まず、環境全体から信号を獲得する(ステップS302)。

【0038】センサ部10、補助センサ部40の各センサは、画像解析部23、音声解析部33、信号解析部43を用いて、ステップS302で得た信号を解析する(ステップS303)。

【0039】データ/制御信号通信手段71~74を通じて、解析結果を出力し(ステップS304)、さらに、解析結果に応じて、他のセンサに対し、信号の獲得の制御に関する指示を出す(ステップS305)。

【0040】センサ部10、補助センサ部40の各センサは、データ/制御信号通信手段71~74を通じて、センサ部10、補助センサ部40のうちの他のセンサと、総合解析部50の解析結果、および他のセンサから受けた信号の獲得制御に関する指示を獲得する(ステップS306)。

【0041】次に、データ/制御信号通信手段71~74を通じて、他のセンサ、総合解析部50から信号の獲得の制御に関する指示の有無を調べる(ステップS307)。

【0042】信号の獲得の制御に関する指示があれば(ステップS307:Y)、他のセンサ、総合解析部50および過去の解析結果を参考した場合、当該指示に従って良いかを判断し(ステップS308)、当該指示に従って良いと判断した場合は(ステップS308:Y)、各センサにおいて撮影装置21、録音装置31、補助センサ装置41を用いて、当該指示に従って環境の信号を獲得する(ステップS309)。

【0043】ステップS308において当該指示に従って良いと判断しなかった場合は(ステップS308:N)、または、ステップS307において信号の獲得の制御に関する指示がない場合(ステップS307:N)、他のセンサの解析結果、総合解析部50の結果、自分のこれまでの解析結果をもとに、環境中において、次に信号を獲得する対象の範囲および獲得する精度を決定し、その決定に従って環境の信号を獲得する(ステップS310)。

【0044】次に、獲得した信号を、他のセンサ、総合解析部50の解析結果、過去の解析結果を参考に解析す

る(ステップS311)。

【0045】データ/制御信号通信手段71~74を通じて、解析結果を出力し(ステップS312)、さらに、解析結果に応じて、他のセンサに対し、信号の獲得の制御に関する指示を出す(ステップS313)。

【0046】ステップS305からステップS313を繰り返す。

【0047】総合解析部50は、次の図4のフローチャートに示した流れにより処理を実行する。

【0048】総合解析部50は、データ/制御信号通信手段75を通じて、アプリケーション60からセンサ部10、補助センサ部40の各センサにおける、信号の獲得の制御に関する指示の有無を調べ(ステップS401)、指示があれば(ステップS401:Y)、その指示を各センサに与える(ステップS402)。指示がない場合は、S403にすすむ。

【0049】総合解析部50は、データ/制御信号通信手段71~74を通じて、各センサから信号処理結果を得て、さらに、データ/制御信号通信手段75を通じて、アプリケーション60から実行結果を獲得する(ステップS403)。

【0050】総合解析部50は、すべての結果のうち、複数の間で最も矛盾の少ない結果を求める(ステップS404)。

【0051】総合解析部50は、データ/制御信号通信手段75を通じて、解析結果をアプリケーション60に提供する(ステップS405)。また、解析結果に応じて、データ/制御信号通信手段71~74を通じて、各センサに信号の獲得制御に関する指示を出す(ステップS406)。

【0052】ステップS401からステップS406を繰り返す。

【0053】以上、本発明の複数のセンサによるヒューマンインタフェースシステムによれば、センサ部10、補助センサ部40の各センサは、相互に解析結果や信号獲得制御に関する指示のやり取りを行い、さらに、総合解析部50による解析結果や信号獲得制御に関する指示のやり取りを行い、センサ間の矛盾を調べつつフィードバック処理を実行し、信号を得る範囲および精度を確実に決定し、複数のセンサ全体で整合性のとれた出力を生成することができる。

【0054】(実施形態2)本発明の実施形態2の複数のセンサによるヒューマンインタフェースシステムは、実施形態1に示した構成を変形した構成としたものである。実施形態1で示した図1の構成は、センサとして、センサ部10の画像センサ20、音声センサ30、補助センサ部40を備えた構成であったが、そのバリエーションとして、図5(a)~(c)の構成がある。図5(a)は、図1の構成から補助センサ部40を省略したもの、図5(b)は、図1の構成から音声センサ30を



省略したもの、図 5 (c) は、図 1 の構成から画像センサ 20 を省略したものとなっている。

【0055】図 5 (a) は、補助センサ部 40 を必要としない場合、画像センサ 20 と音声センサ 30 の間で、データ／制御信号通信手段 71 を通じて、相互に信号処理結果を得て、さらに、データ／制御信号通信手段 73 を通じて、総合解析部 50 から解析結果を獲得し、また、データ／制御信号通信手段 75 を通じてアプリケーション 50 から得た制御指示内容を総合解析部 50 を介して得て、すべての結果のうち、複数の間で最も矛盾の少ない結果を求めるものである。

【0056】図 5 (b) は、画像情報だけでアプリケーションを操作する場合、音声センサ 30 を省略した構成とし、画像センサ 20 と補助センサ部 40 の間で、データ／制御信号通信手段 72 を通じて、相互に信号処理結果を得て、さらに、データ／制御信号通信手段 73、74 を通じて、総合解析部 50 から実行結果を獲得し、また、データ／制御信号通信手段 75 を通じてアプリケーション 60 から得た制御指示内容を総合解析部 50 を介して得て、すべての結果のうち、複数の間で最も矛盾の少ない結果を求めるものである。

【0057】図 5 (c) は、音声情報だけでアプリケーションを操作する場合、画像センサ 20 を省略した構成とし、音声センサ 30 と補助センサ部 40 の間で、データ／制御信号通信手段 72 を通じて、相互に信号処理結果を得て、さらに、データ／制御信号通信手段 73、74 を通じて、総合解析部 50 から実行結果を獲得し、また、データ／制御信号通信手段 75 を通じてアプリケーション 60 から得た制御指示内容を総合解析部 50 を介して得て、すべての結果のうち、複数の間で最も矛盾の少ない結果を求めるものである。

【0058】何れの場合でも、実施形態 1 と同じく、センサ間で処理結果と制御が行き交い、それらの結果を総合して解析することにより、出力を生成し、次のステップで環境中から獲得する信号の範囲と精度を決定する。

【0059】(実施形態 3) 本発明の実施形態 3 の複数のセンサによるヒューマンインタフェースシステムは、実施形態 1 に示した基本構成を、人の音声入力による操作指示、身振りによる操作指示を受け付けるヒューマンインタフェースシステムに適用したものであって、解析結果や信号獲得制御に関する指示のやり取りを行い、センサ間の矛盾を調べつつ複数のセンサ全体で整合性のとれた出力を生成するとともに、検出対象である人の音声入力による操作指示内容、身振りによる操作指示内容を解析して受け付けるシステムである。

【0060】図 6 は、本発明の実施形態 3 にかかる複数のセンサによるヒューマンインタフェースシステムの装置構成概略を示すブロック図である。図 6 に示すように、本発明の実施形態 3 の構成は、センサ部 10a、補助センサ部 40a、総合解析部 50a、アプリケーション

60 を備えている。センサ部 10a は、画像センサ 20a、音声センサ 30a を備えている。なお、検出対象は人、人の声、人の身振りである。

【0061】ここで、画像センサ 20a は、人の画像データを得る機能に加え、人の動きから身振りによる操作指示内容を解析するジェスチャ解析機能を備えている。また、音声センサ 30a は、人の音声データを得る機能に加え、人の音声による操作指示内容を解析する音声解析機能を備えている。また、補助センサ部 40a は、画像センサ 20a、音声センサ 30a の処理に対する有用な信号を得るものであり、例えば、人の位置および人がマネキンやラジオなどの静止物でなく、動いているものであるかを検出するために有用なセンサとして、超音波センサが挙げられ、人がマネキンやラジオなどの無生物ではなく、生物であることを検出するために有用なセンサとして、赤外線センサ（サーモグラフィを生成するセンサ）などが挙げられる。

【0062】また、この構成例では、各データ／制御信号通信手段 71～74、75 は、1 つの通信路 76 を共用する形で形成されている。

【0063】各センサの構成と処理内容をさらに詳しく説明する。

【0064】画像センサ 20a は、撮影装置 21、撮影装置制御部 22、画像解析部 23a、通信部 24 を備えている。画像解析部 23a は、人物検出部 231、人物映像記憶部 232、動作認識部 233、動作パターン記憶部 234 を備えている。

【0065】撮影装置 21 は、カメラ等、映像を獲得する一又は複数の装置であって、撮影範囲の情景を画像として獲得する。動的に撮影方向を変えられるものとする。

【0066】撮影制御部 22 は、画像解析部 23a の画像解析による人物検出結果、音声センサ 30a の音声による音源検出結果、補助センサ 40a による人物位置検出結果に基づき、画像情報の獲得を指向する方向、位置、光学パラメータを決定した上で、撮影装置 21 を制御する。なお、総合解析部 50a やアプリケーション 60 によって撮影制御の指示を受ける場合もある。

【0067】画像解析部 23a は、画像を解析する部分であるが、本実施形態 3 では、画像解析により人物検出とジェスチャ入力によるコマンド認識を行う。

【0068】人物検出部 231 は、画像情報から人物に相当する映像パターンを探索する部分である。撮影装置 21 により獲得された画像情報から、人物映像記憶部 232 に蓄積された画像パターンを参考にして人物画像パターンを探索し、当該人物画像パターンの画像上での位置と撮影装置 21 の位置から人物位置を決定し、画像上での見え方に応じて人物の向きを決定する。

【0069】人物映像記憶部 232 は、探索対象となる画像パターンを蓄積する部分であり、ここでは人間だと



認識できる映像パターンを記憶している。

【0070】動作認識部233は、人物検出部231で検出された人物画像データの動き（ジェスチャ）を解析して、身振り（ジェスチャ）によるコマンドを認識する部分である。動作認識にあたり、動作パターン記憶部234のデータとのマッチングを行って、その人物がジェスチャにより入力したコマンドを認識する。

【0071】動作パターン記憶部234は、コマンドとそのときの人間の動作パターンの映像を対応づけて記憶する部分である。

【0072】次に、画像センサ20aにおける、音声センサ30a、補助センサ40a、総合解析部50aからのデータ、制御指示のフィードバックを受けた動作認識処理について説明する。

【0073】画像情報と音声情報に基づく動作認識は、音声センサ30aにより検出された人物音声を発する音源位置情報をもとに、当該人物の所定時間の画像情報の変化に基づき人間の動作を認識する。その際、動作パターン記憶部234の情報を参考する。

【0074】次に、画像情報と音声情報と補助センサ40aによる人物検出情報に基づく動作認識は、音声センサ30aにより検出された人物音声を発する音源位置情報と、補助センサ40aにより検出された人物位置情報をもとに、当該人物の所定時間の画像情報の変化に基づき人間の動作を認識する。その際、動作パターン記憶部234の情報を参考する。

【0075】画像情報と音声情報と補助センサ40aによる人物検出情報と総合解析部50aの解析結果とアプリケーション60の制御指示に基づく動作認識は、音声センサ30aにより検出された人物音声を発する音源位置情報と、補助センサ40aにより検出された人物位置情報と、総合解析部50aによる各センサ間で最も矛盾の少ない解析情報と、アプリケーション60からの制御指示内容をもとに、当該人物の所定時間の画像情報の変化に基づき人間の動作を認識する。その際、動作パターン記憶部234の情報を参考する。

【0076】次に、音声センサ30aは、録音装置31、録音装置制御部32、音声解析部33a、通信部24を備えている。音声解析部33aは、音源検出部331、人物音声記憶部332、音声認識部333、音声パターン記憶部334を備えている。

【0077】録音装置31は、マイク等、音声を獲得する一又は複数の装置であり、音声を音声情報として獲得する。動的に指向性を設定できるものとする。

【0078】録音装置制御部32は、画像センサ20aによる人物検出結果、音声解析部33aによる音源検出結果、補助センサ40aによる人物位置検出結果に基づき、音声情報を獲得する方向、位置、音響パラメータを決定した上で、録音装置31を制御する。なお、総合解析部50aやアプリケーション60によって撮影制御の

指示を受ける場合もある。

【0079】音声解析部33aは、音声を解析する部分であるが、本実施形態3では、音声解析により人物検出と音声入力によるコマンド認識を行う。

【0080】音源検出部331は、音声情報から人物に相当する音声パターンを探索する部分である。音声情報と人物音声記憶部332に記憶された人物に相当する音響パターンとを比較して人物音声パターンを持つ音源を探し出し、録音装置31の位置と人物に相当する音響パターンの強度に基づき、人間の発声に相当する音源方向を決定する。

【0081】人物音声記憶部332は、探索対象となる音声パターンを蓄積する部分であり、ここでは人間が発声する音響パターンを記憶している。

【0082】音声認識部333は、音源検出部331で検出された音声入力を解析して、音声入力によるコマンドを認識する部分である。音声認識にあたり、音声パターン記憶部334のデータとのマッチングを行って、その人物が音声により入力したコマンドを認識する。

【0083】音声パターン記憶部334は、コマンドとそのときの人間の動作パターンの音声データを対応づけて記憶する部分である。

【0084】次に、音声センサ30aにおける、画像センサ20a、補助センサ40a、総合解析部50aからのデータ、制御指示のフィードバックを受けた音声認識処理について説明する。

【0085】画像情報と音声情報に基づく音声認識は、画像センサ20aにより検出された人物の映像を有する物体位置情報をもとに、当該人物の所定時間の音声情報に基づき人間の音声を認識する。その際、音声パターン記憶部334の情報を参考する。

【0086】次に、画像情報と音声情報と補助センサ40aによる人物検出情報に基づく音声認識は、画像センサ20aにより検出された人物映像を有する物体位置情報と、補助センサ40aにより検出された人物位置情報をもとに、当該人物の所定時間の音声情報に基づき人間の音声を認識する。その際、音声パターン記憶部334の情報を参考する。

【0087】画像情報と音声情報と補助センサ40aによる人物検出情報と総合解析部50aの解析結果とアプリケーション60の制御指示に基づく音声認識は、画像センサ20aにより検出された人物映像を有する物体位置情報と、補助センサ40aにより検出された人物位置情報と、総合解析部50aによる各センサ間で最も矛盾の少ない解析情報と、アプリケーション60からの制御指示内容をもとに、当該人物の所定時間の音声情報に基づき人間の音声を認識する。その際、音声パターン記憶部334の情報を参考する。

【0088】次に、補助センサ40aは、補助センサ装置41、補助センサ装置制御部42、人物位置検出部4

3a、通信部44を備えている。

【0089】補助センサ装置41は、超音波センサ、赤外センサ、レンジセンサなど、画像、音声以外の情報のセンサであり、環境の動きの変化を観測する一又は複数のセンサを備え、人物位置を検出するに有効な情報を獲得する。動的にその探索範囲・精度を設定できるものとする。

【0090】補助センサ装置制御部42は、画像センサ20aによる人物検出結果、音声解析部33aによる音源検出結果、補助センサ部40aによる人物位置検出結果に基づき、補助センサ部により信号を獲得する方向、位置、特徴パラメータを決定した上で、補助センサ装置41を制御する。なお、総合解析部50aやアプリケーション60によって撮影制御の指示を受ける場合もある。

【0091】人物位置検出部43aは、信号解析部43の一形態であり、補助センサ部40aの出力に基づき、人物の位置を検出する部分である。例えば、補助センサ装置41が超音波センサであれば、反射波の解析により、物体位置を検出し、赤外線センサであれば、人の体温程度の熱放射を持つ物体位置を検出する。

【0092】次に、補助センサ部40aにおける、画像センサ20a、音声センサ30a、総合解析部50aからのデータ、制御指示のフィードバックを受けた人物位置検出処理について説明する。

【0093】画像情報と音声情報に基づく人物位置検出は、画像センサ20aにより検出された人物の映像を有する物体位置情報と音声センサ30aにより検出された音源位置情報をもとに、当該人物の所定時間の補助センサ部40a自身により検出された信号に基づき人間の位置を検出する。

【0094】画像情報と音声情報と補助センサ部40aによる人物検出情報と総合解析部50aの解析結果とアプリケーション60の制御指示に基づく音声認識は、画像センサ20aにより検出された人物映像を有する物体位置情報と、音声センサ30aにより検出された音源位置情報と、補助センサ部40a自身により検出された人物位置情報と、総合解析部50aによる各センサ間で矛盾のない解析情報と、アプリケーション60からの制御指示内容をもとに、当該人物の所定時間の補助センサ部40a自身により検出された信号に基づき人間の位置を検出する。

【0095】総合解析部50aは、人物情報記憶部51、人物識別部52、人物状態認識部53、コマンド辞書記憶部54、コマンド理解部55、出力部56、通信部57を備えている。

【0096】人物情報記憶部51は、検出対象となる特定の人物の映像パターンと音声パターンを含む各種情報を記憶する部分である。

【0097】人物識別部52は、画像センサ20a、音

声センサ30a、補助センサ部40aが検出した人物が誰であるかを識別する部分である。この人物識別部52の識別処理において、画像センサ20aの識別結果、音声センサ30aの識別結果、補助センサ部40aの識別結果と、人物情報記憶部51の情報を総合的に判断して、相互に矛盾した結果が得られている場合は、各センサにおける識別結果の何れかまたは全てが正しくないと判断できる。

【0098】人物状態認識部53は、補助センサ部40aの情報をを用いて算出した人物位置と、画像情報に基づく人物検出結果と、音声情報に基づく音源検出結果に基づき、人物の位置、向き、姿勢を決定する。

【0099】コマンド辞書記憶部54は、このコマンド辞書記憶部54は2種類の情報を記憶する。一つは、アプリケーションが受け付ける各コマンドと、動作パターン記憶部234および音声パターン記憶部334中のコマンドの関連を表すリンクであり、もう一つは、アプリケーションがそのときどきで受理可能なコマンドのリスト（一覧）である。後者は、アプリケーション60の実行中、アプリケーション60によって刻々と書き換えられる。

【0100】コマンド理解部55は、人物識別部52で人物を識別できた場合、画像センサの動作認識部233の動作認識結果と音声認識部333の音声認識結果に基づき、その人物が発信したコマンドを理解する。その際、アプリケーションから提供されるコマンド辞書記憶部54の情報を参考にする。

【0101】出力部56は、人物識別結果と人物状態認識結果とコマンド理解の結果をアプリケーション60に提供する。

【0102】次に、処理の流れを示す。

【0103】図7は、本実施形態3のヒューマンインタフェースシステム全体の処理の流れを示すフローチャートである。なお、各ステップの詳細は追って説明する。

【0104】まず、画像センサ20a、音声センサ30a、補助センサ部40aの各センサと、総合解析部50を並行して起動する（ステップS701）。

【0105】各センサは、撮影装置21、録音装置31、補助センサ装置41を用いて、まず、環境全体から信号を獲得する（ステップS702）。

【0106】画像センサ20a、音声センサ30a、補助センサ部40aの各センサは、画像解析部23a、音声解析部33a、信号解析部43を用いて、ステップS702で得た信号を解析し、解析結果を出力する（ステップS703）。

【0107】画像センサ20a、音声センサ30a、補助センサ部40aの各センサは、データ／制御信号通信手段71～74を通じて、画像センサ20a、音声センサ30a、補助センサ部40aのうちの他のセンサと、総合解析部50の解析結果を獲得する（ステップS70

4)。

【0108】次に、得られた解析結果を参考に、補助センサ部40aを制御して情報を獲得し、補助センサ部40aにおいてその情報に基づき人物位置検出処理を実行する。まず、現在得られている信号で人物位置検出処理を実行できるかを調べ(ステップS705)、実行できる場合(ステップS705:Y)、現在得られている信号を用いて、人物位置検出処理を実行し、当該結果を出力する(ステップS706)。

【0109】ステップS705で現在得られている信号で人物位置検出処理を実行できないと判断した場合(ステップS706:N)、および、実行できると判断してステップS706で人物位置検出処理を実行した場合ともに、現在得られている信号に人物位置検出処理に必要な情報が十分にあったか否かを調べ(ステップS707)、不足している場合(ステップS707:N)、不足している情報を獲得できるように、補助センサ装置41、撮影装置21、録音装置31の次の信号獲得の制御を、補助センサ装置制御部42、撮影装置制御部22、録音装置制御部32に与える(ステップS708)。

【0110】次に、補助センサ部40aの出力結果と、音声センサ30aの音声処理結果と、それまでの画像処理結果を参考に、画像センサ20aは、撮影装置21を制御して画像を獲得し、その色情報を調べ人物位置の決定、および、人間の動作の認識を行う。まず、現在得られている信号で人物動作認識処理を実行できるかを調べ(ステップS709)、実行できる場合(ステップS709:Y)、現在得られている信号を用いて、人物動作認識処理を実行し、当該結果を出力する(ステップS710)。その際、あらかじめ登録しておいた人物映像記憶部232の映像データや動作パターン記憶部234の動作パターンの情報を参考する。

【0111】ステップS709で現在得られている信号で人物動作認識処理を実行できないと判断した場合(ステップS709:N)、および、実行できると判断してステップS710で人物動作認識処理を実行した場合ともに、現在得られている信号に人物動作認識処理に必要な情報が十分にあったか否かを調べ(ステップS711)、不足している場合(ステップS711:N)、不足している情報を獲得できるように、補助センサ装置41、撮影装置21、録音装置31の次の信号獲得の制御を、補助センサ装置制御部42、撮影装置制御部22、録音装置制御部32に与える(ステップS712)。

【0112】次に、補助センサ部40aの出力結果と、画像センサ20aの画像処理結果と、それまでの音声処理結果を参考に、音声センサ30aは、録音装置31を制御して音声を獲得し、その波長情報を調べ音源方向の決定、および、人間の音声の認識を行う。まず、現在得

られている信号で人物音声認識処理を実行できるかを調べ(ステップS713)、実行できる場合(ステップS713:Y)、現在得られている信号を用いて、人物音声認識処理を実行し、当該結果を出力する(ステップS714)。その際、あらかじめ登録しておいた人物音声記憶部332の音声データや音声パターン記憶部334の音声パターンの情報を参考する。

【0113】ステップS713で現在得られている信号で人物動作認識処理を実行できないと判断した場合(ステップS713:N)、および、実行できると判断してステップS714で人物音声認識処理を実行した場合ともに、現在得られている信号に人物音声認識処理に必要な情報が十分にあったか否かを調べ(ステップS715)、不足している場合(ステップS715:N)、不足している情報を獲得できるように、補助センサ装置41、撮影装置21、録音装置31の次の信号獲得の制御を、補助センサ装置制御部42、撮影装置制御部22、録音装置制御部32に与える(ステップS716)。

【0114】次に、総合解析部50aにおいて、画像情報、音声情報、補助センサの情報に基づき、人物の状態認識、人物識別、コマンド理解を行う(ステップS717)。人物識別においては人物識別部52は、あらかじめ人物情報記憶部51に登録しておいた特定の人物の映像情報や音声情報を参考にして人物を識別し、コマンド理解においてはコマンド理解部55は、コマンド辞書記憶部54の情報を参考する。

【0115】ステップS717で得たコマンド理解の結果は、出力部56を通してアプリケーション60に出力される(ステップS718)。なお、アプリケーション60に対して、画像情報、音声情報、補助センサの解析結果も併せて出力しても良い。

【0116】信号処理を続行する場合(ステップS719:Y)、ステップS705へ戻り、信号処理を繰り返す。

【0117】以上が、本実施形態3による処理全体の流れである。

【0118】次に、各センサにおける処理や、各センサが相互に連動し合い、結果を参照し合う処理についてさらに詳しく説明する。

【0119】まず、補助センサ部40aにおける処理を詳しく述べる。

【0120】補助センサ装置41として、超音波センサを用いる場合は、少なくとも1つ以上の超音波発信素子と受信素子を用意し、人間からの反射波を受信した際の受信信号の振幅情報や位相情報に基づき3次元空間内の物体配置情報を計算する。

【0121】超音波センサの振幅情報と位相情報を用いると静止物体以外の動物体を感知することができる。したがって、環境空間中の動物体を検出することにより人

間の位置を検出する。

【0122】補助センサ装置41として赤外センサを用いる場合は、環境の温度分布を画像の形態にて獲得する。人間の体温の部分に相当し、かつ、人間の大きさに相当する領域を見つけることにより、人間がいる場所を特定する。

【0123】補助センサ装置41としてレンジセンサを用いた場合、ある時点とその次の時点の3次元情報の差を求めることにより動く物体がわかり、その大きさが通常の人間に相当するものを見つけることで人間がいる場所を特定する。

【0124】補助センサ部40aの情報を、撮影装置、録音装置が利用する一例を挙げる。超音波センサを用いた場合、空間内の動きのある物体の位置情報を撮影装置および録音装置に提供することができる。位置情報が既知となることで、撮影装置は広い空間内のごく限られた領域のみで人物検出を行えば良く、かかる処理時間を大幅に減らすことが可能となる。また、位置情報が既知となることで、録音装置自身が音源位置を検出しなくとも良くなる。

【0125】次に、補助センサ部40aが、撮影装置、録音装置からの情報を利用する一例を挙げる。撮影装置21および録音装置31は、超音波センサからの位置情報に関わらず、各装置の機能によって人物位置や音源位置を検出した場合、超音波センサはその位置情報をもとに位置近傍に走査し、実際に動きのある物体（人物）の有無を検知する。画像装置は人物以外の部分を人物位置と認識したり、録音装置はTVやオーディオといった人物でない音源からの音を検出する可能性があり、こうした誤検出を防止する効果がある。

【0126】また、補助センサ部40aが、自身で検出する空間内の動きのある物体の位置情報を利用する一例を挙げる。補助センサ部40aは、既知となった動きのある物体の位置情報を利用することで、その物体近傍のみに限定して走査することで高速化し、物体の移動に追従するように走査位置をダイナミックに変更することが可能で、リアルタイム性の高いアクティブな物体追跡センサとしての効果がある。

【0127】次に、画像センサ20aを中心とした処理について詳しく述べる。

【0128】画像センサ20aは、2台以上の複数の撮影装置、または、1台の撮影装置であっても撮影方向および撮影位置およびピント、ズーム、絞りを始めとした光学パラメータの制御のすべてまたは一部を変更できるものまたはどれも変更できないものを用いる。

【0129】画像情報に基づく人物検出をする際、撮影装置21から人間がいる場所への方向のほか、距離も求めたい場合は、2台以上の複数の撮影装置を用いるか、1台の撮影装置であっても撮影位置を変更できるか、撮影位置を変更できなかったとしてもピント、ズーム、絞

りを始めとした光学パラメータを変更できるものを使用する。その場合、何れの装置設定であってもアプリケーションがコマンドを受けつけたい範囲の人間に対しては、異なる撮影位置、または、異なる光学パラメータによって2枚以上の画像を獲得できるように配置する。以上の条件が満たされれば、魚眼レンズを用いた撮影装置、CMOS形式によって画像の任意の画素を取り出せる撮影装置、可視光以外の色を獲得できる撮影装置を用いることもできる。

【0130】撮影する方向は、補助センサ部40aを用いて人物を検出した結果と、音声センサ30aの音声情報を用いて音源方向を検出した結果と、画像情報に基づき人物位置を検出したそれまでの結果を参考する。すなわち、ある時点の前で人物位置を検出できていて、かつ、それが音源方向と一致し、実際に人間であったと判断した場合は、その付近を観察できるように撮影方向を定める。しかし、直前に人間の存在を判断した方向と、補助センサ部40aが示した人間が存在する方向が異なっていた場合、また、直前に人間の確認ができなかった場合は、補助センサ部40aによる人物検出結果の付近を観察できるように撮影方向を定める。

【0131】ただし、人間が移動できる速さを越えて、撮影方向を変更することはしない。例えば、0.1秒後に5m離れた地点に撮影の向きを変更することはしない。直前の結果と補助センサ部40aの結果が異なる場合は、直前の結果に従った方向を再度、撮影する。また、人間の存在を確認できなかった場合も、人間の存在を確認できた最後の撮影方向と補助センサ部40aが指示する方向を比較し、人間の移動速度を超えて撮影方向が変化しているならば、最後に撮影した方向を維持する。

【0132】撮影におけるピント、ズーム、絞りを始めとした光学パラメータは、直前に検出した人間の位置の近辺を撮影するように調整する。補助センサ部40aが人物までの距離などを出力できる場合は、方向の制御と同様の仕組みで参考にする。

【0133】画像による人物検出、音声による音源検出、補助センサ40aによる人物位置検出で、遮断物などの影響により人物や音源の検出をするに至らなかった場合は、各検出過程から撮影すべき方向や位置が指定されるので、指定された方向や位置を観察できるように撮影装置21を制御する。

【0134】次に、人物映像記憶部232における人物映像パターン情報の登録と、動作パターン記憶部234における動作パターンの登録について説明する。

【0135】人物検出部231の人物検出処理や動作認識部233の動作認識処理に先立ち、人物映像記憶部232に対する人物映像パターン情報の登録と、動作パターン記憶部234に対する動作パターン登録を行う。

【0136】人物映像記憶部232には、人間と認識で

き、さらに、その人間の位置、向きを特定できる人物映像パターン情報を記憶する。人間を撮影する方向は、人間の向きの特定に必要な画像を獲得できる方向とする。これは、あらかじめ配置する撮影装置21の位置、アプリケーション60の出力を表示する表示装置およびアプリケーションが操作を許可する表示装置に対する人間の向きの範囲に基づく。人間を撮影する際の画像の解像度（「大きさ」と読み替えてもよい。画素数で換算する面積に相当する。）は、人間の向きを特定するのに必要な解像度とする。人間の向きを顔の向きによって定める場合は、顔の輪郭、目、鼻、口などの位置を判別できる解像度の画像を記憶する。人間の向きを目の向きによって定める場合は、目を判別でき、かつ、視線を判別できるように白目、黒目の位置を判別できるように撮影する。

【0137】以上のように、事前に人間を撮影しておく画像は、撮影装置と人間の向き、距離に応じて複数の画像を撮影する場合がある。また、アプリケーションの要求項目に応じて、表情（笑い、泣きなど）の異なる画像、化粧、ヒゲ、メガネの有無の画像などを撮影する場合もありうる。

【0138】以上の情報を復元できるのであれば、記憶する画像のデータ量を圧縮しておいてもよい。また、人間の向き、位置を特定できるのであれば、人間の姿全体ではなく、顔、目等の部分のみを記憶しておいてもよい。人間の姿全体を圧縮した情報と、特徴のある部分の圧縮していない情報を合わせるなど、以上の情報を組み合わせることもできる。

【0139】動作パターン記憶部234には、アプリケーション60がコマンドと認知するのに必要な人間の動作を記憶する。また、手、足、頭、ひじなど、人間の主要な関節部分の肌の座標変化を動作パターンとして記憶することもできる。

【0140】次に、画像センサ20aにおける画像情報に基づく人物検出処理について述べる。人物検出処理は以下のように行う。

【0141】人物検出で用いる画像センサ20aで取り込まれる画像情報は、補助センサ部40aの出力、音声センサ30aによる音源検出結果、および、直前の画像センサ20aによる人物検出結果を参考にして撮影されたものとする。入力された画像中、顔に相当する部分領域の探索には画素の色情報と位置情報を用いる。そして、ピンホールカメラなどのカメラモデルにしたがって撮影装置の位置と画像上の検出位置から人間が存在する方向を計算する。撮影装置が複数あり、複数の画像上で検出できた場合は、三角測量の原理に基づき、人間が存在する方向のみならず位置も計算する。

【0142】人間の顔をあらわす画像情報は、人物映像記憶部232に登録されたものを用いる。後述するように人間の顔は、アプリケーションの仕様にしたがって正面から撮影した映像のほか、横から撮影した画像を含

む。その場合は、登録されているすべての向きから撮影した見え方の顔を検出する。

【0143】なお、直前まで人間の存在を確認できていたにも関わらずできなくなった場合は、各センサに感度や精度を上げて情報の確認を行うように指示する。すなわち、録音装置にはボリュームやサンプリング間隔を上げるように、補助センサ部40aには環境のスキャンニングの間隔を狭めるように、撮影装置にはズームなどにより解像度を上げて撮影するように指示する。それでも、人物を検出できないと判断した場合は、撮影した方向に人間がいなかったと判断する。

【0144】次に、画像センサ20aにおける画像情報と音声情報に基づく動作認識は以下に行われる。動作認識で用いる画像情報は、補助センサ部40aの出力、音声センサ20aによる音源検出結果、および、画像センサ20aの直前の人物検出結果を参考にして撮影されたものとする。前述の動作パターン記憶部234に記憶された情報と同様に、入力画像から情報を抽出し、その変化を、動作パターン記憶部234に記憶されている動作パターンと照合する。入力画像中における動作と動作パターン中の動作の一致は、動画像の連続する各画像における情報を記憶した動作パターンにおけるそれらとの差をとり、動画像全体を通した累積値により判断する。もしくは、入力画像における変化を、離散コサイン変換、フーリエ変換などによって近似し、同じく近似した動作パターン中の動作と、変換された係数の比較によって一致を判断する。

【0145】なお、人間の存在を確認できたにも関わらず動作によるコマンド理解ができなかった場合は、各センサに感度や精度を上げて情報の確認を行うように指示する。すなわち、録音装置にはボリュームやサンプリング間隔を上げるように、補助センサ40aには環境のスキャンニングの間隔を狭めるように、撮影装置にはズームなどにより解像度を上げて撮影するように指示する。

【0146】次に、音声センサ30aを中心とした処理について詳しく説明する。

【0147】音声センサ30aは、2台以上の複数の録音装置、または、1台の録音装置をアプリケーションがコマンドを受け付けようとする範囲の人間の音声を獲得できるように配置する。

【0148】録音装置の出力は、補助センサ部40aを用いて人物を検出した結果と、画像情報を用いて人物位置を検出した結果と、音声情報に基づき音源方向を検出したそれまでの結果を参考にした、特定の方向の音声に関する信号とする。すなわち、ある時点の前で人物の発声に相当する音声を検出できていて、かつ、それが画像から判断した人間が存在する方向と一致し、実際に人間であったと判断した場合は、その付近を指向した音声信号に関する情報を出力する。しかし、直前に人間の存在を判断した方向と、補助センサ40aが示した人間が存



在する方向が異なっていた場合、また、直前に人間の確認ができなかった場合は、補助センサ部40aによる人物検出結果の付近を指向した音声信号に関する情報を出力する。

【0149】なお、画像による人物検出、音声による音源検出、補助センサ40aによる人物位置検出で、遮断物などの影響により人物や音源の検出をするに至らなかった場合は、各検出過程から録音装置31が録音すべき指向方向が指定されるので、その方向を指向した音声情報を出力する。

【0150】次に、音源検出処理を述べる。音源検出で用いる音声情報は、補助センサ部40aの出力、画像センサ20aによる人物検出結果、および、直前の音源検出結果を参考にして特定の方向を指向して出力された音声情報とする。音声情報の周波数特性を調べ、後述する人物音声として記憶された情報の特徴との一致を判断することにより人物音声の有無を判断する。人物音声がないと判断した場合は、音声情報の獲得を指向した方向に人間がいなかったと判断する。

【0151】複数の録音装置31をある程度離して配置した場合は、音声録音された時間の差に基づき、音源位置を特定する。また、複数箇所にそれぞれ複数の録音装置を配置した場合は、それぞれの箇所において音源方向を特定し、それぞれの箇所の録音装置群から計算された音源方向の重なるところを求めることにより、音源位置を求める。直前まで人間の音声を確認できていたにも関わらず音源を検出できなくなった場合は、各センサに感度や精度を上げて情報の確認を行うように指示する。すなわち、録音装置にはボリュームやサンプリング間隔を上げるように、補助センサ40aには環境のスキャンニングの間隔を狭めるように、撮影装置にはズームなどにより解像度を上げて撮影するように指示する音声情報と画像情報に基づく音声認識において、当該音声認識で用いる画像情報は、補助センサ部40aの出力、音源検出結果、および、直前の人物検出結果を参考にして撮影されたものとする。なお、人間の音声を検出できたにも関わらず音声によるコマンド理解ができなかった場合は、各センサに感度や精度を上げて情報の確認を行うように指示する。すなわち、録音装置にはボリュームやサンプリング間隔を上げるように、補助センサ40aには環境のスキャンニングの間隔を狭めるように、撮影装置にはズームなどにより解像度を上げて撮影するように指示する。

【0152】次に、総合解析部50aを中心にした処理について詳しく説明する。

【0153】人物識別部52による人物識別処理や、人物状態認識部53における人物状態認識処理に先立ち、人物情報記憶部51の登録情報やコマンド辞書記憶部54の登録情報を更新する。

【0154】人物情報記憶部51には、映像情報、音声

情報とも、アプリケーションが操作を許可する特定の個人名または複数名を識別するのに必要な情報を記憶する。映像情報は、人物を特定するのに必要な、各人を撮影した映像を記憶する。事前に人間を撮影しておく画像は、撮影装置と人間の向き、距離に応じて複数の画像を撮影する場合がある。また、アプリケーションの要求項目に応じて、表情（笑い、泣きなど）の異なる画像、化粧、ヒゲ、メガネの有無の画像などを撮影する場合もありうる。人物の特定に必要な画像を復元できるのであれば、記憶する画像のデータ量を圧縮しておいてもよい。また、個人差を確認するのに必要な、顔のみ、または、さらに分解して目、口、鼻などの部分だけを記憶してもよい。また、以上の情報を組み合わせたり、または、復元可能であれば圧縮の上、組み合わせて記憶することもできる。

【0155】音声情報は、人物を特定するのに必要な、各人が発生した音響を記憶する。各人の発声は個別に記憶する。アプリケーションが受け付けるコマンドを記憶するか、または、アプリケーションが受け付けるコマンドを復元するのに必要な発声を記憶する。後者の場合は、コマンドを分解した音素や単語を録音することになる。記憶の際、録音された音声をそのまま記憶するのではなくコマンドまたは音素または単語ごとの周波数特性または周波数特性の変化を記憶することもできる。以上の情報を組み合わせたり、または、復元可能であれば圧縮の上、組み合わせて記憶することもできる。

【0156】コマンド辞書記憶部54には2種類の情報を記憶する。一つは、アプリケーションが受け付ける各コマンドと、動作パターン記憶部234および音声パターン記憶部334中のコマンドの関連を表すリンクであり、もう一つは、アプリケーションがそのときどきで受理可能なコマンドのリスト（一覧）である。後者のコマンド辞書は、アプリケーションの進行にしたがって逐次、アプリケーションが書き換える、したがって、動作認識結果と音声認識結果に基づくコマンド理解処理は、そのときどきの処理において、絶えずコマンド辞書を参照する。

【0157】次に、総合解析部50aの人物識別部52による人物識別について述べる。人物識別部52は、画像センサ20aの人物検出と音声センサ30aの音源検出において、それぞれ、人物映像の確認、人物音声の検出ができ、かつ、人物検出結果と音源検出結果でそれぞれ検出した位置が重なる場合のみ、人物を観察できたと判断する。後述の人物識別、コマンド理解においては、人物を確認できた場合の画像情報、音声情報のみ有効とする。人物識別部52が人物を観察できたと判断した場合、人間の位置、方向を、補助センサ40aによる人物位置検出結果、画像情報による人物検出結果、および、音声情報による音源検出結果の重なる位置、方向として出力する。姿勢は、画像情報による人物検出結果に従

う。

【0158】人物識別部52は、画像情報および音声情報を、人物情報記憶部51にあらかじめ登録された特定の人物の画像情報および音声情報と照合することにより判断する。照合に先立ち、撮影装置21で獲得した画像情報や録音装置31で獲得した音声情報は、特定の人物を撮影した映像情報や特定の人物の発声を録音した音声情報を、人物情報記憶部51に格納する際に行う方法と同じ方法で加工する。

【0159】例えば、画像情報と音声情報が有効で、かつ、画像情報から判断した人物と音声情報から判断した人物が一致している場合は、その人物を人物識別結果とし、そうでない場合は、画像情報から判断した結果に付けた得点と、音声情報から判断した結果に付けた得点の大きい方の結果を人物識別結果とする。両者の得点を比較する際、画像情報と音声情報間のバランスを定める重み係数を用意し、その重みを加味した上で判断してもよい。

【0160】さらに、画像情報から判断した人物と音声情報から判断した人物の履歴を参考にしてもよい。画像情報および音声情報から人物を判断した際、それぞれ得点を付けて複数の候補を選出する。それを先の方法にしたがい組合せ、画像情報および音声情報の両方を参考した得点付きの複数の候補を求める。これらの得点を人物候補ごとに累積し、累積値があらかじめ定めたある値を超えた時点で、その人物を認識したと判断する。いったん人物が確定すると、人物検出、音声検出が一定時間途絶えるまではこれを保持する。人物検出、音声検出が一定時間途絶えた場合は、この履歴を抹消し、再度、特定の累積を開始する。人物の履歴を参考する場合で、人物状態認識部53では人間の存在が確認できなかったが画像情報もしくは音声情報のどちらか一方が得られた場合、その画像もしくは音声情報に対して人物の識別を行い、得られた得点を低めた上でその得点を累積する。ただし、人物の特定の判断、すなわち、累積値があらかじめ定めたある値を超えたかの判断は、人物状態認識部53で人物の存在が確認された場合のみ行う。

【0161】次に、人物状態認識部53における人物状態認識処理を説明する。ここで、人物の状態とは、人物の位置、方向、姿勢のことである。補助センサ部40aからは人間がいる方向と補助センサ40aの機能により距離がわかる。画像センサ20aの画像情報からは、人間の存在を確認できた場合に限り方向と距離もわかり、さらに顔の向きに関する情報が登録されている場合は姿勢（顔の向き）もわかる。音声センサ30aの音声情報からは、人間の音声を確認できた場合に限り人間の音声が聞こえてくる方向がわかり、環境の複数箇所にそれぞれ複数の録音装置を配置した場合は、環境内での人間の位置がわかる。

【0162】次に、動作認識結果と音声認識結果に基づ

くコマンド理解について説明する。

【0163】人物識別で人物を識別できていると判断されている間の動作認識結果と音声認識結果のそれぞれに対し、それぞれの結果から得られたコマンドについて、コマンド辞書記憶部54に格納されたアプリケーション60のコマンドとの対応を調べ、動作から判断したコマンドと音声から判断したコマンドが同じアプリケーションのコマンドに対応していて、かつ、そのコマンドが現在のアプリケーション60が受け付けることのできるコマンドのリストに入っていた場合は、同コマンドを理解したと判断する。なお、人物識別で人物を識別できていると判断されている状態で、動作認識による結果と音声認識による結果のどちらか一方しか得られていない場合であれば、その結果から得られたコマンドが現在のアプリケーション60が受け付けることのできるコマンドのリストに入っている場合は、同コマンドを理解したと判断する。

【0164】以上のように理解したコマンド内容をアプリケーションへ提供する。人物状態認識部53によって認識した人物の位置、方向、姿勢に関する情報と、人物識別部52で特定した人物を指し示す情報と、コマンド理解部55にて理解したコマンドに関する情報を、アプリケーション60から問い合わせがあった場合に返答する。他の扱いとしては、アプリケーション60からの問い合わせを待たず、これらの情報が得られたら即座にアプリケーション60に通信部57などを用いてアプリケーション60に通知する処理でも良く、必要な情報が揃った時点でアプリケーション60に通知するものでも良い。

【0165】次に、総合解析部50aにおけるコマンド辞書の生成処理について説明する。アプリケーション60の開始に先立ち、コマンド辞書記憶部54中、アプリケーション60が受け付ける各コマンドと、動作パターン記憶部234および音声パターン記憶部334中のコマンドの関連を表すリンクをコマンド辞書記憶部54に格納する。また、コマンド辞書記憶部54のアプリケーション60が受理可能なコマンドのリストは、アプリケーション60が開始した後、アプリケーション60が状況に応じて設定する。このリストを空にすることもできる。

【0166】（実施形態4）実施形態4の構成例を図8に示す。

【0167】実施形態4にかかる図8の構成例では、不特定人物の動作コマンド、音声コマンドおよび人物の状態を理解する。実施形態3で説明した図6の構成例に比べて総合解析部50bの構成が異なるものとなっている。この構成は総合解析部50bが、人物情報記憶部51、人物識別部52を持っていない。また、コマンド理解部55は、人物状態認識部53の結果を、人物識別部52を介さず直接受け取る構成となっている。



【0168】この構成によれば、人物の識別処理を特に行わず、入力されたすべての人物に関する情報を処理し、不特定人物の動作コマンド、音声コマンドおよび人物の状態を理解する。

【0169】（実施形態5）本発明の複数のセンサによるヒューマンインタフェースシステムは、上記に説明した構成を実現する処理ステップを記述したプログラムをコンピュータ読み取り可能な記録媒体に記録して提供することにより、各種コンピュータを用いて構築することができる。本発明の複数のセンサによるヒューマンインタフェースシステムを実現する処理ステップを備えたプログラムを記録した記録媒体は、図9に図示した記録媒体の例に示すように、CD-ROM1002やフレキシブルディスク1003等の可搬型記録媒体1001だけでなく、ネットワーク上にある記録装置内の記録媒体1000や、コンピュータのハードディスクやRAM等の記録媒体1005のいずれであっても良く、プログラム実行時には、プログラムはコンピュータ1004上にローディングされ、主メモリ上で実行される。

【0170】本発明の複数のセンサによるヒューマンインタフェースシステムについてさらに以下の項を開示する。

【0171】（付記1） 検出対象の範囲と検出精度を決定し、当該検出精度により当該検出対象から特定の検出信号を獲得するセンサとして、検出信号が異なる少なくとも2種類以上のセンサと、前記各センサで検出された結果間に相互に矛盾がないか否かを調べ、各センサに対する制御情報を生成する総合解析部と、前記各センサによる検出結果を利用するアプリケーションと、前記各センサ相互間、前記各センサと前記総合解析部間、前記総合解析部と前記アプリケーション間においてデータと制御情報を通信し合う通信手段を備え、前記各センサが、他のセンサから得た前記信号検出結果または制御情報、前記総合解析部から得た前記制御情報のいずれかまたはその組み合わせを利用して、次の信号獲得時における検出対象の範囲と検出精度を決定し、前記各センサが、前記通信手段を介して、他のセンサに対して、自身の信号検出結果と、他のセンサが次の信号獲得時における検出対象の範囲と検出精度の決定に利用する制御情報を出力し、前記総合解析部が、前記通信手段を介して、前記各センサが次の信号獲得時における検出対象の範囲と検出精度の決定に利用する制御情報を出力することを特徴とする複数のセンサによるヒューマンインタフェースシステム（1）。

【0172】（付記2） 前記検出対象が人であり、前記センサが少なくとも画像センサと音声センサを含み、前記画像センサの検出信号が人の画像認識情報であり、前記画像センサが、画像認識結果に基づいて検出対象の動作を解釈し、ジェスチャ入力によるコマンド入力を実行する動作認識部を備え、前記音声センサの検出信号が

人の音声認識情報であり、前記音声センサが、音声認識結果に基づいて検出対象の音声を実行し、音声入力によるコマンド入力を実行する音声認識部を備えた付記1に記載の複数のセンサによるヒューマンインタフェースシステム（2）。

【0173】（付記3） 前記検出対象が人であり、前記センサが少なくとも画像センサと補助センサを含み、前記画像センサの検出信号が人の画像認識情報であり、前記画像センサが、画像認識結果に基づいて検出対象の動作を解釈し、ジェスチャ入力によるコマンド入力を実行する動作認識部を備え、前記補助センサの検出信号が人の位置情報を検出するために有効な情報である付記1に記載の複数のセンサによるヒューマンインタフェースシステム（3）。

【0174】（付記4） 前記検出対象が人であり、前記センサが少なくとも音声センサと補助センサを含み、前記音声センサの検出信号が人の音声認識情報であり、前記音声センサが、音声認識結果に基づいて検出対象の音声を実行し、音声入力によるコマンド入力を実行する音声認識部を備え、前記補助センサの検出信号が人の位置情報を検出するために有効な情報である付記1に記載の複数のセンサによるヒューマンインタフェースシステム（4）。

【0175】（付記5） 前記検出対象が人であり、前記センサが少なくとも画像センサと音声センサと補助センサを含み、前記画像センサの検出信号が人の画像認識情報であり、前記画像センサが、画像認識結果に基づいて検出対象の動作を解釈し、ジェスチャ入力によるコマンド入力を実行する動作認識部を備え、前記音声センサの検出信号が人の音声認識情報であり、前記音声センサが、音声認識結果に基づいて検出対象の音声を実行し、音声入力によるコマンド入力を実行する音声認識部を備え、前記補助センサの検出信号が人の位置情報を検出するために有効な情報である付記1に記載の複数のセンサによるヒューマンインタフェースシステム（5）。

【0176】（付記6） 前記補助センサとして、超音波センサ、赤外線センサ、レンジセンサのいずれかまたはそれらの組み合わせを用いた付記3～5のいずれかに記載の複数のセンサによるヒューマンインタフェースシステム。

【0177】（付記7） 検出対象の範囲と検出精度を決定し、当該検出精度により当該検出対象から特定の検出信号を獲得するセンサの処理として、検出信号が異なる少なくとも2種類以上のセンサの処理と、前記各センサで検出された結果間に相互に矛盾がないか否かを調べ、各センサの処理に対する制御情報を生成する総合解析部と、前記各センサ処理による検出結果を利用するアプリケーション処理と、前記各センサ処理相互間、前記各センサ処理と前記総合解析部処理間、前記総合解析部処理と前記アプリケーション処理間においてデータと制御

情報を通信し合う通信処理を備え、前記各センサ処理が、他のセンサ処理から得た前記信号検出結果または制御情報、前記総合解析処理から得た前記制御情報のいずれかまたはその組み合わせを利用して、次の信号獲得時における検出対象の範囲と検出精度を決定し、前記各センサ処理が、前記通信手段を介して、他のセンサ処理に対して、自身の信号検出結果と、他のセンサ処理が次の信号獲得時における検出対象の範囲と検出精度の決定に利用する制御情報を出力し、前記総合解析処理が、前記通信手段を介して、前記各センサ処理が次の信号獲得時における検出対象の範囲と検出精度の決定に利用する制御情報を出力することを特徴とする複数のセンサによるヒューマンインタフェース方法。

【0178】（付記8） 前記検出対象が人であり、前記センサ処理が少なくとも画像センサ処理と音声センサ処理を含み、前記画像センサ処理の検出信号が人の画像認識情報であり、前記画像センサ処理が、画像認識結果に基づいて検出対象の動作を解釈し、ジェスチャ入力によるコマンド入力を認識する動作認識処理を備え、前記音声センサ処理の検出信号が人の音声認識情報であり、前記音声センサ処理が、音声認識結果に基づいて検出対象の音声を解釈し、音声入力によるコマンド入力を認識する音声認識処理を備えた付記7に記載の複数のセンサによるヒューマンインタフェース処理方法。

【0179】（付記9） 前記検出対象が人であり、前記センサ処理が少なくとも画像センサ処理と補助センサ処理を含み、前記画像センサ処理の検出信号が人の画像認識情報であり、前記画像センサ処理が、画像認識結果に基づいて検出対象の動作を解釈し、ジェスチャ入力によるコマンド入力を認識する動作認識処理を備え、前記補助センサ処理の検出信号が人の位置情報を検出するために有効な情報である付記7に記載の複数のセンサによるヒューマンインタフェース処理方法。

【0180】（付記10） 前記検出対象が人であり、前記センサ処理が少なくとも音声センサ処理と補助センサ処理を含み、前記音声センサ処理の検出信号が人の音声認識情報であり、前記音声センサ処理が、音声認識結果に基づいて検出対象の音声を解釈し、音声入力によるコマンド入力を認識する音声認識処理を備え、前記補助センサ処理の検出信号が人の位置情報を検出するために有効な情報である付記7に記載の複数のセンサによるヒューマンインタフェース処理方法。

【0181】（付記11） 前記検出対象が人であり、前記センサ処理が少なくとも画像センサ処理と音声センサ処理と補助センサ処理を含み、前記画像センサ処理の検出信号が人の画像認識情報であり、前記画像センサ処理が、画像認識結果に基づいて検出対象の動作を解釈し、ジェスチャ入力によるコマンド入力を認識する動作認識処理を備え、前記音声センサ処理の検出信号が人の音声認識情報であり、前記音声センサ処理が、音声認識

結果に基づいて検出対象の音声を解釈し、音声入力によるコマンド入力を認識する音声認識処理を備え、前記補助センサ処理の検出信号が人の位置情報を検出するために有効な情報である付記7に記載の複数のセンサによるヒューマンインタフェース処理方法。

【0182】（付記12） 前記補助センサ処理として、超音波センサ処理、赤外センサ処理、レンジセンサ処理のいずれかまたはそれらの組み合わせを用いた付記9～11のいずれかに記載の複数のセンサによるヒューマンインタフェース処理方法。

【0183】

【発明の効果】本発明の複数のセンサによるヒューマンインタフェースシステムによれば、画像情報と音声情報を同時に扱うことによって、実際には声がしない人の顔を映した写真やマネキン人形を人間として認識してしまうことや、人間ではない音響機器のスピーカーから流れ出た音声情報を認識してしまうことがなくなり、マンマシンインタフェースの誤動作が減る。同様に、画像情報と音声情報が処理結果を互いに利用することで、声がしないところの映像を処理対象にしたり、人間がいないところの音声を処理対象することがなくなり、処理を誤る確率が減る。また、処理対象データが減ることから、処理速度が向上する。

【0184】本発明の複数のセンサによるヒューマンインタフェースシステムによれば、補助センサを用いて、画像情報と音声情報では扱えない情報を獲得することにより、画像や音声を処理における探索範囲を狭め、処理速度を向上できる。また、補助センサの情報を利用することにより、撮影装置や録音装置の配置数を減らすことができる。

【図面の簡単な説明】

【図1】 本発明の実施形態1にかかる複数のセンサによるヒューマンインタフェースシステムの装置構成概略を示す図

【図2】 図1の基本構成を適用し、図1の画像センサ20、音声センサ30、補助センサ40構成例を詳しく示した図

【図3】 本発明の実施形態1の複数のセンサによるヒューマンインタフェースシステムの信号処理の流れを示したフローチャート

【図4】 本発明の実施形態1にかかる総合解析部50の信号処理の流れを示したフローチャート

【図5】 本発明の実施形態2にかかる、実施形態1に示した構成を変形した構成例を示す図

【図6】 本発明の実施形態3にかかる複数のセンサによるヒューマンインタフェースシステムの装置構成概略を示すブロック図

【図7】 本発明の実施形態3のヒューマンインタフェースシステム全体の処理の流れを示すフローチャート

【図8】 本発明の実施形態4にかかる複数のセンサに

よるヒューマンインタフェースシステムの装置構成概略を示すブロック図

【図9】 本発明の実施形態5にかかる複数のセンサによるヒューマンインタフェースシステムを実現する処理プログラムを記録した記録媒体を示した図

【図10】 従来におけるセンサを用いたインタフェースの例を示す図

【図11】 従来におけるセンサを用いたインタフェースの他の例を示す図

【符号の説明】

10 センサ部

20, 20a 画像センサ

21 撮影装置

22 撮影装置制御部

23, 23a 画像解析部

24 通信部

30, 30a 音声センサ

31 録音装置

32 録音装置制御部

33, 33a 音声解析部

34 通信部

40, 40a 補助センサ部

41 補助センサ装置

42 補助センサ装置制御部

43 信号解析部

43a 人物位置検出部

44 通信部

50, 50a, 50b 総合解析部

51 人物情報記憶部

52 人物識別部

53 人物状態認識部

54 コマンド辞書記憶部

55 コマンド理解部

56 出力部

57 通信部

60, 60a アプリケーション

61 通信部

71, 72, 73, 74, 75 データ/制御信号通信手段

76 通信路

231 人物検出部

232 人物映像記憶部

233 動作認識部

234 動作パターン記憶部

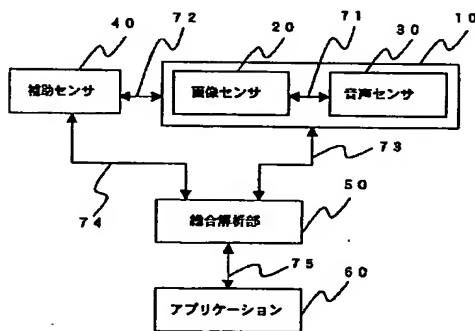
331 音源検出部

332 人物音声記憶部

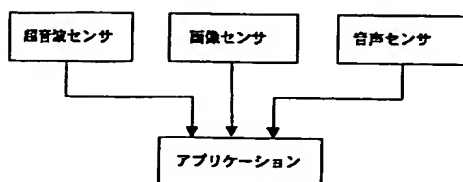
333 音声認識部

334 音声パターン記憶部

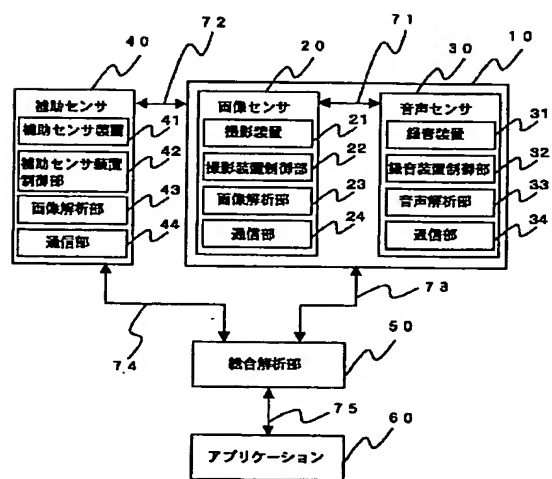
【図1】



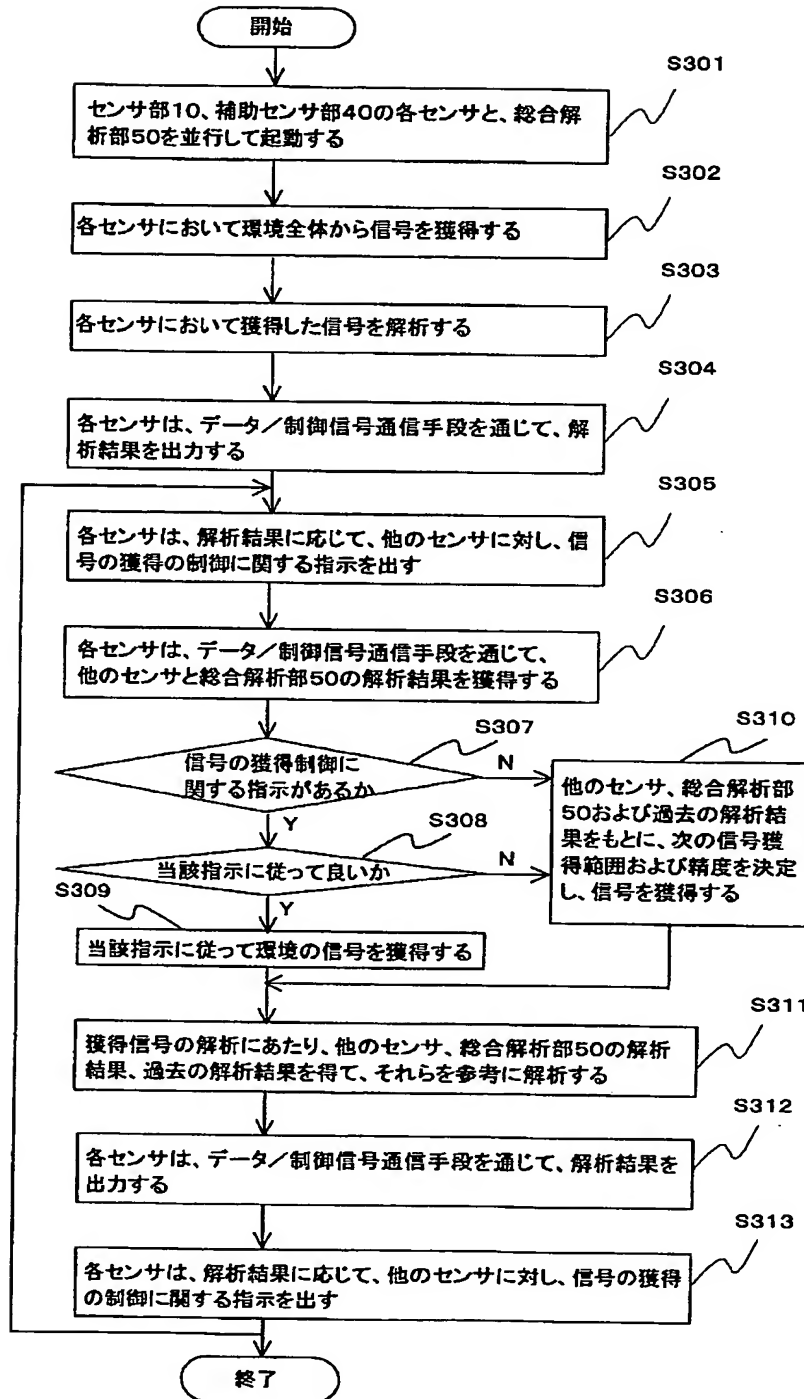
【図10】



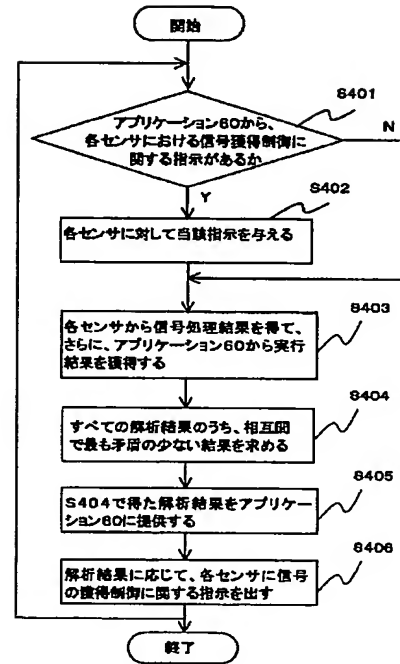
【図2】



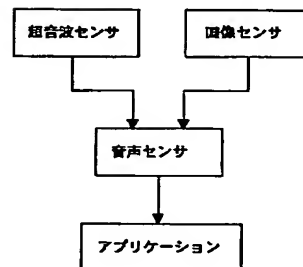
【図3】



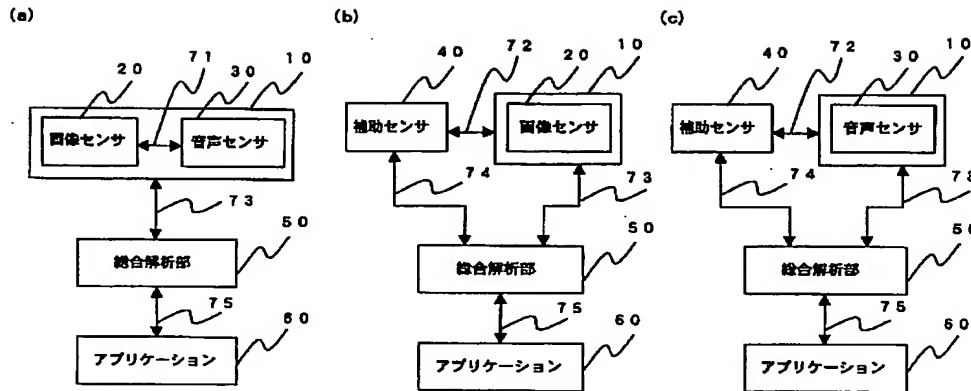
【図4】



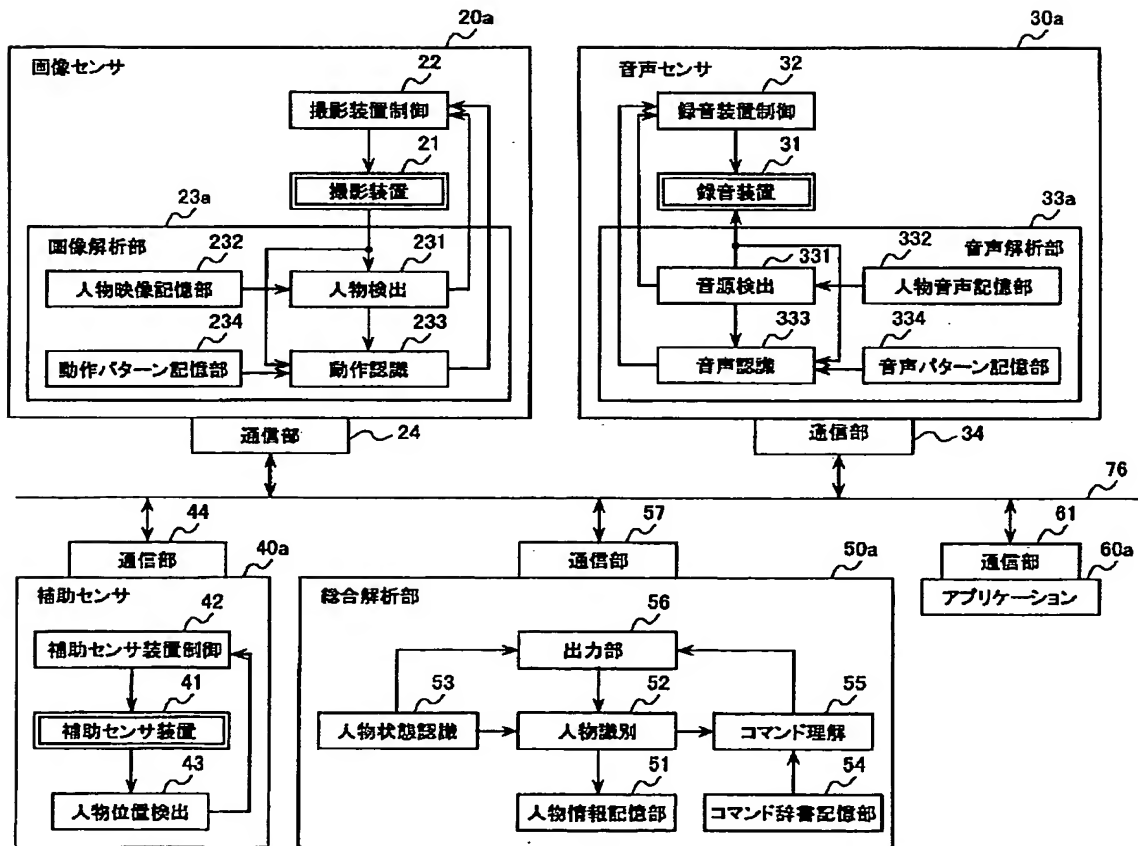
【図11】



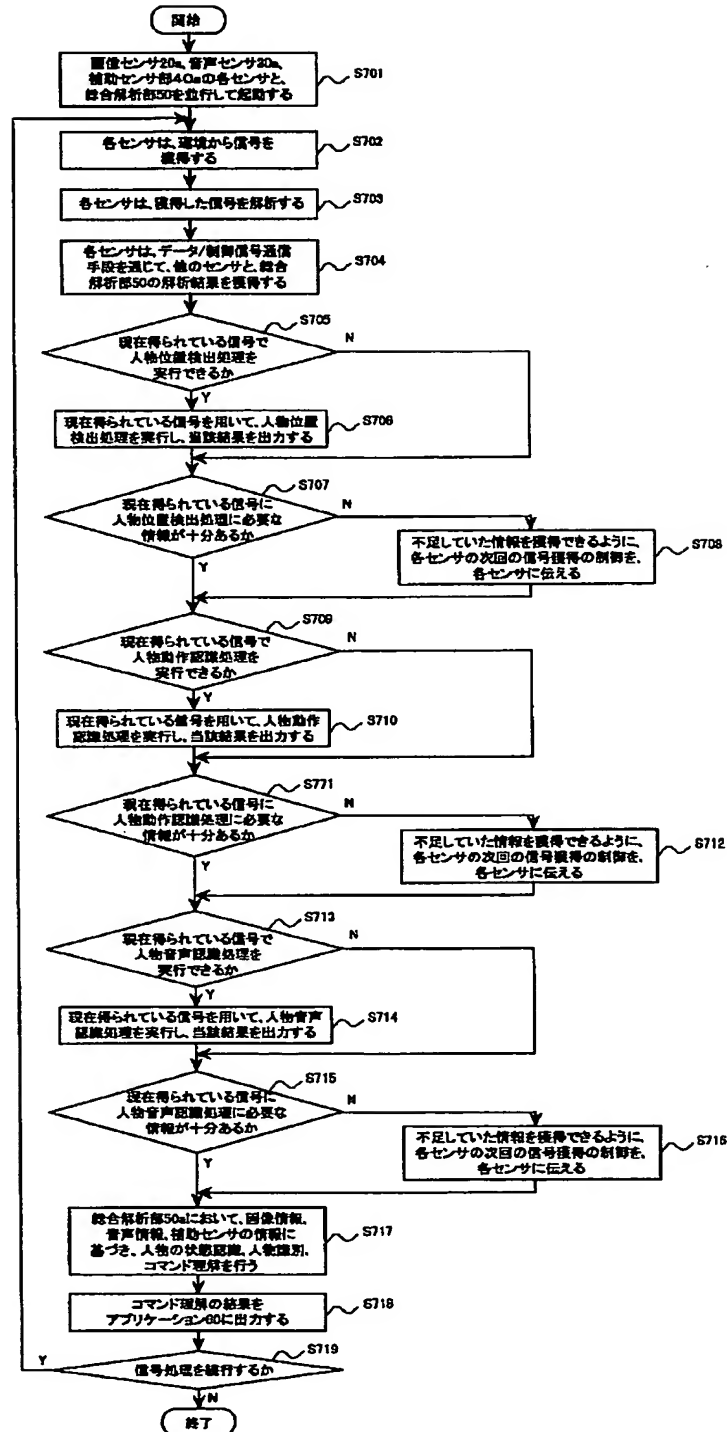
【図5】



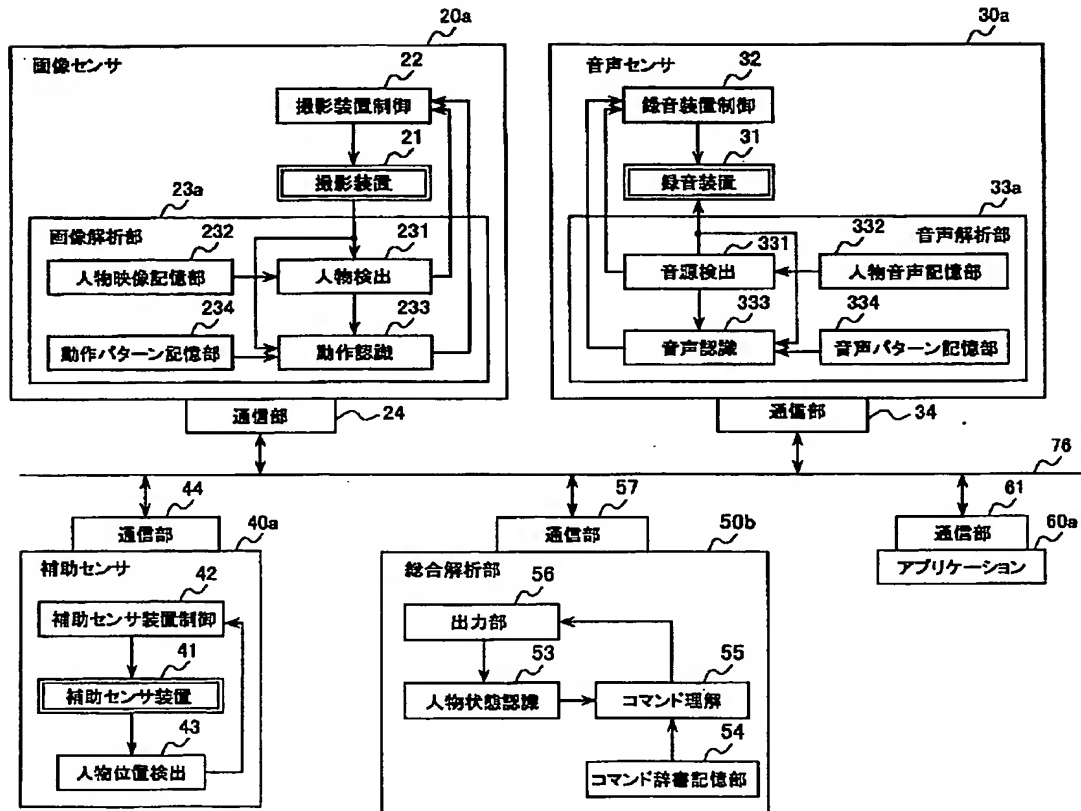
【図6】



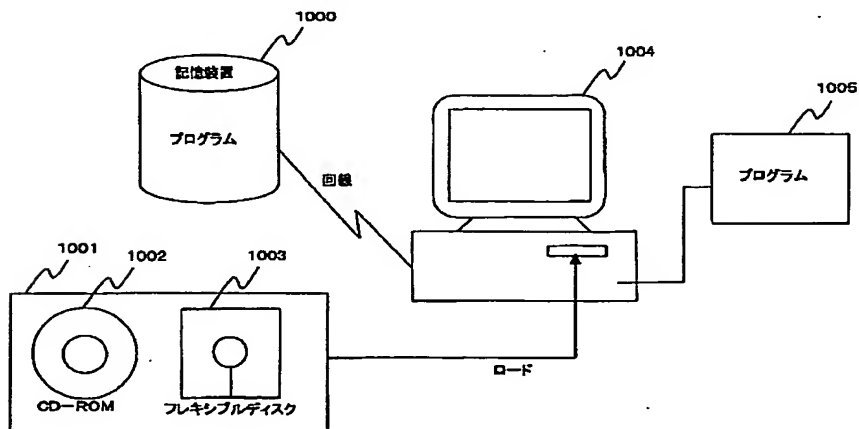
【図7】



【図 8】



【図 9】





## フロントページの続き

(51) Int. Cl. 7	識別記号	F I	キーワード (参考)
G 1 0 L 15/00		G 1 0 L 3/00	5 5 1 F
(72) 発明者 野田 拓也		F ターム (参考) 5D015 KK01	
神奈川県川崎市中原区上小田中 4 丁目 1 番			5E501 AA02 AB06 AC37 BA01 CB14
1 号 富士通株式会社内			CB15 EA21
(72) 発明者 渡辺 一宏			5L096 BA16 BA18 BA20 CA05 DA02
神奈川県川崎市中原区上小田中 4 丁目 1 番			HA02
1 号 富士通株式会社内			